

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

IN RE APPLICATION OF: Kenichiro KOBAYASHI

GAU:

SERIAL NO: New Application

EXAMINER:

FILED: Herewith

FOR: SINGING VOICE SYNTHESIZING METHOD AND APPARATUS, PROGRAM, RECORDING
MEDIUM AND ROBOT APPARATUS

REQUEST FOR PRIORITY

COMMISSIONER FOR PATENTS
ALEXANDRIA, VIRGINIA 22313

SIR:

☐ Full benefit of the filing date of U.S. Application Serial Number , filed , is claimed pursuant to the provisions of 35 U.S.C. §120.

☐ Full benefit of the filing date(s) of U.S. Provisional Application(s) is claimed pursuant to the provisions of 35 U.S.C. §119(e): Application No. Date Filed

☒ Applicants claim any right to priority from any earlier filed applications to which they may be entitled pursuant to the provisions of 35 U.S.C. §119, as noted below.

In the matter of the above-identified application for patent, notice is hereby given that the applicants claim as priority:

COUNTRY

APPLICATION NUMBER

MONTH/DAY/YEAR

Japan

2003-079151

March 20, 2003

Certified copies of the corresponding Convention Application(s)

☒ are submitted herewith

☐ will be submitted prior to payment of the Final Fee

☐ were filed in prior application Serial No. filed

☐ were submitted to the International Bureau in PCT Application Number

Receipt of the certified copies by the International Bureau in a timely manner under PCT Rule 17.1(a) has been acknowledged as evidenced by the attached PCT/IB/304.

☐ (A) Application Serial No.(s) were filed in prior application Serial No. filed ; and

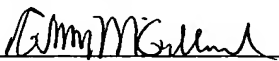
☐ (B) Application Serial No.(s)

☐ are submitted herewith

☐ will be submitted prior to payment of the Final Fee

Respectfully Submitted,

OBLON, SPIVAK, McCLELLAND,
MAIER & NEUSTADT, P.C.


Bradley D. Lytle

Registration No. 40,073

Customer Number

22850

Tel. (703) 413-3000
Fax. (703) 413-2220
(OSMMN 05/03)

C. Irvin McClelland
Registration Number 21,124

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日
Date of Application: 2 0 0 3 年 3 月 2 0 日

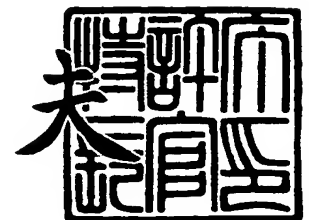
出 願 番 号
Application Number: 特 願 2 0 0 3 - 0 7 9 1 5 1
[ST. 10/C]: [J P 2 0 0 3 - 0 7 9 1 5 1]

出 願 人
Applicant(s): ソニー株式会社

2 0 0 4 年 1 月 1 3 日

特許庁長官
Commissioner,
Japan Patent Office

今 井 康 夫



【書類名】 特許願

【整理番号】 0390001805

【提出日】 平成15年 3月20日

【あて先】 特許庁長官 殿

【国際特許分類】 G06F 17/20
G10L 13/00
G10H 7/00

【発明者】

【住所又は居所】 東京都品川区北品川 6 丁目 7 番 3 5 号 ソニー株式会社
内

【氏名】 小林 賢一郎

【特許出願人】

【識別番号】 000002185

【氏名又は名称】 ソニー株式会社

【代理人】

【識別番号】 100067736

【弁理士】

【氏名又は名称】 小池 晃

【選任した代理人】

【識別番号】 100086335

【弁理士】

【氏名又は名称】 田村 榮一

【選任した代理人】

【識別番号】 100096677

【弁理士】

【氏名又は名称】 伊賀 誠司

【手数料の表示】

【予納台帳番号】 019530

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9707387

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 歌声合成方法、歌声合成装置、プログラム及び記録媒体、並びにロボット装置

【特許請求の範囲】

【請求項 1】 演奏データを音の高さ、長さ、歌詞の音楽情報として解析する解析工程と、

解析された音楽情報に基づき、歌声を生成する歌声生成工程と、

上記演奏データに基づき、上記歌声の伴奏としての楽音を生成する楽音生成工程とを有することを特徴とする歌声合成方法。

【請求項 2】 上記演奏データは M I D I ファイルの演奏データであることを特徴とする請求項 1 記載の歌声合成方法。

【請求項 3】 上記楽音生成工程は上記歌声の対象とした演奏データに係る楽音をミュートすることを特徴とする請求項 1 記載の歌声合成方法。

【請求項 4】 上記楽音生成工程は上記演奏データのうち、予め指定されたトラックの演奏データに係る楽音をミュートすることを特徴とする請求項 2 記載の歌声合成方法。

【請求項 5】 上記楽音生成工程は上記歌声の対象とした演奏データに係る楽音を上記歌声の音量よりも小さな音量で再生することを特徴とする請求項 1 記載の歌声合成方法。

【請求項 6】 上記歌声と上記楽音の同期を取ってミキシングするミキシング工程をさらに有することを特徴とする請求項 1 記載の歌声合成方法。

【請求項 7】 上記ミキシング工程は上記歌声生成工程からの上記歌声と上記楽音生成工程からの上記楽音をミキシングする際に、それぞれの波形データを予め作成し重ね合わせることによりミキシングをすることを特徴とする請求項 6 記載の歌声合成方法。

【請求項 8】 演奏データを音の高さ、長さ、歌詞の音楽情報として解析する解析手段と、

解析された音楽情報に基づき、歌声を生成する歌声生成手段と、

上記演奏データに基づき、上記歌声の伴奏としての楽音を生成する楽音生成手

段とを有することを特徴とする歌声合成装置。

【請求項 9】 上記演奏データは M I D I ファイルの演奏データであることを特徴とする請求項 8 記載の歌声合成装置。

【請求項 10】 上記楽音生成手段は上記歌声の対象とした演奏データに係る楽音をミュートすることを特徴とする請求項 8 記載の歌声合成装置。

【請求項 11】 上記楽音生成手段は上記歌声の対象とした演奏データに係る楽音を上記歌声の音量よりも小さな音量で再生することを特徴とする請求項 8 記載の歌声合成装置。

【請求項 12】 上記歌声と上記楽音の同期を取ってミキシングするミキシング手段をさらに有することを特徴とする請求項 8 記載の歌声合成装置。

【請求項 13】 所定の処理をコンピュータに実行させるためのプログラムであって、

演奏データを音の高さ、長さ、歌詞の音楽情報として解析する解析工程と、
解析された音楽情報に基づき、歌声を生成する歌声生成工程と、

上記演奏データに基づき、上記歌声の伴奏としての楽音を生成する楽音生成工程とを有することを特徴とするプログラム。

【請求項 14】 上記演奏データは M I D I ファイルの演奏データであることを特徴とする請求項 13 記載のプログラム。

【請求項 15】 上記歌声と上記楽音の同期を取ってミキシングするミキシング工程をさらに有することを特徴とする請求項 13 記載のプログラム。

【請求項 16】 所定の処理をコンピュータに実行させるためのプログラムが記録されたコンピュータ読み取り可能な記録媒体であって、

演奏データを音の高さ、長さ、歌詞の音楽情報として解析する解析工程と、
解析された音楽情報に基づき、歌声を生成する歌声生成工程と、

上記演奏データに基づき、上記歌声の伴奏としての楽音を生成する楽音生成工程とを有することを特徴とするプログラムが記録された記録媒体。

【請求項 17】 上記演奏データは M I D I ファイルの演奏データであることを特徴とする請求項 16 記載の記録媒体。

【請求項 18】 上記歌声と上記楽音の同期を取ってミキシングするミキシン

グ工程をさらに有することを特徴とする請求項16記載の記録媒体。

【請求項19】 供給された入力情報に基づいて動作を行う自律型のロボット装置であって、

入力された演奏データを音の高さ、長さ、歌詞の音楽情報として解析する解析手段と、

解析された音楽情報に基づき、歌声を生成する歌声生成手段と、

上記演奏データに基づき、上記歌声の伴奏としての楽音を生成する楽音生成手段とを有することを特徴とするロボット装置。

【請求項20】 上記演奏データはMIDIファイルの演奏データであることを特徴とする請求項19記載のロボット装置。

【請求項21】 上記歌声と上記楽音の同期を取ってミキシングするミキシング手段をさらに有することを特徴とする請求項20記載のロボット装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、演奏データから歌声を合成する歌声合成方法、歌声合成装置、プログラム及び記録媒体、並びにロボット装置に関する。

【0002】

【従来の技術】

コンピュータ等により、与えられた歌唱データから歌声を生成する技術は特許文献1に代表されるように既に知られている。

【0003】

MIDI (musical instrument digital interface) データは代表的な演奏データであり、事実上の業界標準である。代表的には、MIDI データはMIDI 音源と呼ばれるデジタル音源 (コンピュータ音源や電子楽器音源等のMIDI データにより動作する音源) を制御して楽音を生成するのに使用される。MIDI ファイル (例えば、SMF (standard MIDI file)) には歌詞データを入れることができ、歌詞付きの楽譜の自動作成に利用される。

【0004】

また、MIDI データを歌声又は歌声を構成する音素セグメントのパラメータ表現（特殊データ表現）として利用する試みも特許文献2に代表されるように提案されている。

【0005】

しかし、これらの従来の技術においてはMIDI データのデータ形式の中で歌声を表現しようとしているが、あくまでも楽器をコントロールする感覚でのコントロールであり、MIDI 本来が持っている歌詞データを利用するものではなかった。

【0006】

また、ほかの楽器用に作成されたMIDI データを、修正を加えることなく歌声にすることはできなかった。

【0007】

また、電子メールやホームページを読み上げる音声合成ソフトはソニー(株)の「Simple Speech」をはじめ多くのメーカーから発売されているが、読み上げ方は普通の文章を読み上げるのと同じような口調であった。

【0008】

ところで、電氣的又は磁氣的な作用を用いて人間（生物）の動作に似た運動を行う機械装置を「ロボット」という。我が国においてロボットが普及し始めたのは、1960年代末からであるが、その多くは、工場における生産作業の自動化・無人化等を目的としたマニピュレータや搬送ロボット等の産業用ロボット（Industrial Robot）であった。

【0009】

最近では、人間のパートナーとして生活を支援する、すなわち住環境その他の日常生活上の様々な場面における人的活動を支援する実用ロボットの開発が進められている。このような実用ロボットは、産業用ロボットとは異なり、人間の生活環境の様々な局面において、個々に個性の相違した人間、又は様々な環境への適応方法を自ら学習する能力を備えている。例えば、犬、猫のように4足歩行の動物の身体メカニズムやその動作を模した「ペット型」ロボット、あるいは、2足直立歩行を行う人間等の身体メカニズムや動作をモデルにしてデザインされた

「人間型」又は「人間形」ロボット (Humanoid Robot) 等のロボット装置は、既に実用化されつつある。

【0010】

これらのロボット装置は、産業用ロボットと比較して、エンターテインメント性を重視した様々な動作を行うことができるため、エンターテインメントロボットと呼称される場合もある。また、そのようなロボット装置には、外部からの情報や内部の状態に応じて自律的に動作するものがある。

【0011】

この自律的に動作するロボット装置に用いられる人工知能 (A I : artificial intelligence) は、推論・判断等の知的な機能を人工的に実現したものであり、さらに感情や本能等の機能をも人工的に実現することが試みられている。このような人工知能の外部への表現手段としての視覚的な表現手段や自然言語の表現手段等のうちで、自然言語表現機能の一例として、音声を用いることが挙げられる。

【0012】

【特許文献1】

特許第3233036号公報

【特許文献2】

特開平11-95798号公報

【0013】

【発明が解決しようとする課題】

以上のように従来の歌声合成は特殊な形式のデータを用いていたり、仮にM I D I データを用いていてもその中に埋め込まれている歌詞データを有効に活用できなかったり、ほかの楽器用に作成されたM I D I データを歌い上げたりすることはできなかった。

【0014】

本発明は、このような従来の実情に鑑みて提案されたものであり、例えばM I D I データのような演奏データを活用して歌声を合成することが可能な歌声合成方法及び装置を提供することを目的とする。

【0015】

さらに、本発明の目的は、MIDIデータのような演奏データを活用する際、歌声に使用し、かつ歌声と共にもとの演奏データから楽音も再生可能した歌声合成方法及び装置を提供することである。

【0016】

さらに、本発明の目的は、このような歌声合成機能をコンピュータに実施させるプログラム及び記録媒体を提供することである。

【0017】

さらに、本発明の目的は、このような歌声合成機能を実現するロボット装置を提供することである。

【0018】**【課題を解決するための手段】**

本発明に係る歌声合成方法は、上記目的を達成するため、演奏データを音の高さ、長さ、歌詞の音楽情報として解析する解析工程と、解析された音楽情報に基づき、歌声を生成する歌声生成工程と、上記演奏データに基づき、上記歌声の伴奏としての楽音を生成する楽音生成工程とを有することを特徴とする。

【0019】

また、本発明に係る歌声合成装置は、上記目的を達成するため、演奏データを音の高さ、長さ、歌詞の音楽情報として解析する解析手段と、解析された音楽情報に基づき、歌声を生成する歌声生成手段と、上記演奏データに基づき、上記歌声の伴奏としての楽音を生成する楽音生成手段とを有することを特徴とする。

【0020】

この構成によれば、本発明に係る歌声合成方法及び装置は、演奏データを解析してそれから得られる歌詞や音の高さ、長さ、強さをもとにした音符情報に基づき歌声情報を生成し、その歌声情報をもとに歌声の生成を行うことができる。さらに、演奏データから歌声の伴奏としての楽音を再生することにより、伴奏のもとで歌詞を歌い上げることができる。

【0021】

上記演奏データはMIDIファイル（例えばSMF）の演奏データであること

が好ましい。

【0022】

上記楽音生成工程又は手段は歌声を目立たせるために上記歌声の対象とした演奏データに係る楽音をミュートする（楽音として出力しない）ことが好ましい。

【0023】

あるいは、上記楽音生成工程又は手段は上記歌声の対象とした演奏データに係る楽音を上記歌声の音量よりも小さな音量で再生することにより、カラオケ等におけるメロディガイド機能を果たすことができる。

【0024】

また、上記楽音生成工程又は手段は上記MIDIファイルの演奏データのうち、歌詞の対象等として予め指定されたトラックの演奏データに係る楽音をミュートすることが好ましい。

【0025】

さらに、上記歌声と上記楽音の同期を取ってミキシングするミキシング工程又は手段を設けることが好ましい。ミキシングの方式としては、歌声と楽音のそれぞれの波形データを予め作成し重ね合わせるによりミキシングをすることとし、ミキシング結果を保存してもよい。

【0026】

また、本発明に係るプログラムは、本発明の歌声合成機能をコンピュータに実行させるものであり、本発明に係る記録媒体は、このプログラムが記録されたコンピュータ読み取り可能なものである。

【0027】

さらに、本発明に係るロボット装置は、上記目的を達成するため、供給された入力情報に基づいて動作を行う自律型のロボット装置であって、入力された演奏データを音の高さ、長さ、歌詞の音楽情報として解析する解析手段と、解析された音楽情報に基づき、歌声を生成する歌声生成手段と、上記演奏データに基づき、上記歌声の伴奏としての楽音を生成する楽音生成手段とを有することを特徴とする。これにより、ロボットの持っているエンターテインメント性を格段に向上させることができる。

【0028】

【発明の実施の形態】

以下、本発明を適用した具体的な実施の形態について、図面を参照しながら詳細に説明する。

【0029】

先ず、本実施の形態における歌声合成装置の概略システム構成を図1に示す。ここで、この歌声合成装置は、少なくとも感情モデル、音声合成手段及び発音手段を有する例えばロボット装置に適用することを想定しているが、これに限定されず、各種ロボット装置や、ロボット以外の各種コンピュータ A I (artificial intelligence) 等への適用も可能であることは勿論である。

【0030】

図1において、MIDIデータに代表される演奏データ1を解析する演奏データ解析部2は入力された演奏データ1を解析し演奏データ内にあるトラックやチャンネルの音の高さや長さ、強さを表す楽譜情報4に変換する。

【0031】

図2に楽譜情報4に変換された演奏データ(MIDIデータ)の例を示す。図2において、トラック毎、チャンネル毎にイベントが書かれている。イベントにはノートイベントとコントロールイベントが含まれる。ノートイベントは発生時刻(図中の時間の欄)、高さ、長さ、強さ(velocity)の情報を持つ。したがって、ノートイベントのシーケンスにより音符列又は音列が定義される。コントロールイベントは発生時刻、コントロールのタイプデータ(例えばビブラート、演奏ダイナミクス表現(expression))及びコントロールのコンテンツを示すデータを持つ。例えば、ビブラートの場合、コントロールのコンテンツとして、音の振れの大きさを指示する「深さ」、音の揺れの周期を指示する「幅」、音の揺れの開始タイミング(発音タイミングからの遅れ時間)を指示する「遅れ」の項目を有する。特定のトラック、チャンネルに対するコントロールイベントはそのコントロールタイプについて新たなコントロールイベント(コントロールチェンジ)が発生しない限り、そのトラック、チャンネルの音符列の楽音再生に適用される。さらに、MIDIファイルの演奏データにはトラック単位で歌詞を記入する

ことができる。図2において、上方に示す「あるうひ」はトラック1に記入された歌詞の一部であり、下方に示す「あるうひ」はトラック2に記入された歌詞の一部である。すなわち図2の例は、解析した音楽情報（楽譜情報）の中に歌詞が埋め込まれた例である。

【0032】

なお、図2において、時間は「小節：拍：ティック数」で表され、長さは「ティック数」で表され、強さは「0-127」の数値で表され、高さは440Hzが「A4」で表される。また、ビブラートは、深さ、幅、遅れがそれぞれ「0-64-127」の数値で表される。

【0033】

図1に戻り、変換された楽譜情報4は歌詞付与部5に渡される。歌詞付与部5では楽譜情報4をもとに音符に対応した音の長さ、高さ、強さ、表情などの情報とともにその音に対する歌詞が付与された歌声情報6の生成を行う。

【0034】

図3に歌声情報6の例を示す。図3において、「¥song¥」は歌詞情報の開始を示すタグである。タグ「¥PP, T10673075¥」は10673075 μ secの休みを示し、タグ「¥tdyna 110 649075¥」は先頭から10673075 μ secの全体の強さを示し、タグ「¥fine100¥」はMIDIのファインチューンに相当する高さの微調整を示し、タグ「¥vibrato NRPN__dep=64¥」、[¥vibrato NRPN__del=50¥]、「¥vibrato NRPN__rat=64¥」はそれぞれ、ビブラートの深さ、遅れ、幅を示す。また、タグ「¥dyna 100¥」は音毎の強弱を示し、タグ「¥G4, T288461¥あ」はG4の高さで、長さが288461 μ secの歌詞「あ」を示す。図3の歌声情報は図2に示す楽譜情報（MIDIデータの解析結果）から得られたものである。

【0035】

図3と図2の比較から分かるように、楽器制御用の演奏データ（例えば音符情報）が歌声情報の生成において十分に活用されている。例えば、歌詞「あるうひ」の構成要素「あ」について、「あ」以外の歌唱属性である「あ」の音の発生時

刻、長さ、高さ、強さ等について、楽譜情報（図 2）中のコントロール情報やノートイベント情報に含まれる発生時刻、長さ、高さ、強さ等が直接的に利用され、次の歌詞要素「る」についても楽譜情報中の同じトラック、チャンネルにおける次のノートイベント情報が直接的に利用され、以下同様である。

【0036】

図 1 に戻り、歌声情報 6 は歌声生成部 7 に渡される。歌声生成部 7 は音声合成器（speech synthesizer）を構成する。歌声生成部 7 においては歌声情報 6 をもとに歌声波形 8 の生成を行う。ここで、歌声情報 6 から歌声波形 8 を生成する歌声生成部 7 は例えば図 4 に示すように構成される。

【0037】

図 4 において、歌声韻律生成部 7-1 は歌声情報 6 を歌声韻律データに変換する。波形生成部 7-2 は歌声韻律データを歌声波形 8 に変換する。

【0038】

具体例として、「A4」の高さの歌詞要素「ら」を一定時間伸ばす場合について説明する。ビブラートをかけない場合の歌声韻律データは、以下の表のように表される。

【0039】

【表 1】

[LABEL]		[PITCH]	[VOLUME]	
0	ra	0 50	0	66
1000	aa		39600	57
39600	aa		40100	48
40100	aa		40600	39
40600	aa		41100	30
41100	aa		41600	21
41600	aa		42100	12
42100	aa		42600	3
42600	aa			
43100	a.			

【0 0 4 0】

この表において、[LABEL]は、各音韻の継続時間長を表したものである。すなわち、「r a」という音韻（音素セグメント）は、0サンプルから1 0 0 0サンプルまでの1 0 0 0サンプルの継続時間長であり、「r a」に続く最初の「a a」という音韻は、1 0 0 0サンプルから3 9 6 0 0サンプルまでの3 8 6 0 0サンプルの継続時間長である。また、[P I T C H]は、ピッチ周期を点ピッチで表したものである。すなわち、0サンプル点におけるピッチ周期は5 6サンプルである。ここでは「ら」の高さを変えないので全てのサンプルに渡り5 6サンプルのピッチ周期が適用される。また、[VOLUME]は、各サンプル点での相対的な音量を表したものである。すなわち、デフォルト値を1 0 0 %としたときに、0サンプル点では6 6 %の音量であり、3 9 6 0 0サンプル点では5 7 %の音量である。以下同様にして、4 0 1 0 0サンプル点では4 8 %の音量等が続き4 2 6 0 0サンプル点では3 %の音量となる。これにより「ら」の音声時間が時間の経過と共に減衰することが実現される。

【0 0 4 1】

これに対して、ビブラートをかける場合には、例えば、以下に示すような歌声

韻律データが作成される。

【 0 0 4 2 】

【表 2】

[LABEL]	[PITCH]	[VOLUME]
0 ra	0 50	0 66
1000 aa	1000 50	39600 57
11000 aa	2000 53	40100 48
21000 aa	4009 47	40600 39
31000 aa	6009 53	41100 30
39600 aa	8010 47	41600 21
40100 aa	10010 53	42100 12
40600 aa	12011 47	42600 3
41100 aa	14011 53	
41600 aa	16022 47	
42100 aa	18022 53	
42600 aa	20031 47	
43100 a.	22031 53	
	24042 47	
	26042 53	
	28045 47	
	30045 53	
	32051 47	
	34051 53	
	36062 47	
	38062 53	
	40074 47	
	42074 53	
	43100 50	

【0043】

この表の[PITCH]の欄に示すように、0サンプル点と1000サンプル点におけるピッチ周期は50サンプルで同じであり、この間は音声の高さに変化がないが、それ以降は、2000サンプル点で53サンプルのピッチ周期、4009サンプル点で47サンプルのピッチ周期、6009サンプル点で53のピッチ周期というようにピッチ周期が約4000サンプルの周期(幅)を以て上下(50 ± 3)に振れている。これにより音声の高さの揺れであるビブラートが実現される。この[PITCH]の欄のデータは歌声情報6における対応歌声要素(例えば「ら」)に関する情報、特にノートナンバー(例えばA4)とビブラートコントロールデータ(例えば、タグ「¥vibrato NRPN_dep=64¥」、[¥vibrato NRPN_del=50¥]、「¥vibrato NRPN_rat=64¥」)に基づいて生成される。

【0044】

波形生成部7-2はこのような歌声音韻データに基づき、音素セグメントデータを記憶するデータメモリ(図示せず)から該当するサンプルを読み出して歌声波形8を生成する。すなわち、波形生成部7-2は、データメモリを参照しながら、歌声音韻データに示される音韻系列、ピッチ周期、音量等をもとに、なるべくこれに近い音素セグメントデータを検索してその部分を切り出して並べ、音声波形データを生成する。すなわち、データメモリには、例えば、CV(Consonant, Vowel)や、VCV、CVC等の形で音素セグメントデータが記憶されており、波形生成部7-2は、歌声音韻データに基づいて、必要な音素セグメントデータを接続し、さらに、ポーズ、アクセント、イントネーション等を適切に付加することで、歌声波形8を生成する。なお、歌声情報6から歌声波形8を生成する歌声生成部7については上記の例に限らず、任意の適当な公知の音声合成器を使用できる。

【0045】

図1に戻り、演奏データ1はMIDI音源9に渡され、MIDI音源9は演奏データをもとに楽音の生成を行う。この楽音は伴奏波形10である。

【0046】

歌声波形 8 と伴奏波形 1 0 はともに同期を取りミキシングを行うミキシング部 1 1 に渡される。

【 0 0 4 7 】

ミキシング部 1 1 では、歌声波形 8 と伴奏波形 1 0 との同期を取りそれぞれを重ね合わせて出力波形 3 として再生を行うことにより、演奏データ 1 をもとに伴奏を伴った歌声による音楽再生を行う。

【 0 0 4 8 】

ここで、M I D I 音源 9 での楽音の再生は M I D I 制御部 1 2 により M I D I 制御データ 1 6 に指示されているトラック又はチャンネルに対して、ミュートや音量の調節を行った上で再生が行われる。

【 0 0 4 9 】

M I D I 制御データ 1 6 は、歌詞付与部 5 において歌詞を付与する際にどのトラックに対して歌詞を付与するかを判別、設定するトラック選択部 1 3 において選択されたトラック又はチャンネルの情報も反映され、M I D I 音源 9 からの楽音と歌声生成部 7 が生成する歌声データを同時に再生する際に、歌声の対象となるトラック又はチャンネルに対して自動的にミュート又は音量の調整の処置を施すことができる。

【 0 0 5 0 】

また、これとは別にオペレータの指示により、任意のトラック又はチャンネルに対してもミュート又は音量の調整を施すことができる。

【 0 0 5 1 】

これらの M I D I 制御データ 1 6 は演奏の対象となる M I D I データと例えばファイル名が同じで拡張子が異なるなどの形で関連付けを持って保存することが可能である。

【 0 0 5 2 】

一般に M I D I 音源 9 は再生する楽音を wav 形式などの波形データとして保存することも可能である。ミキシング部 1 1 は歌声データとのミキシングを行う際に、この予め用意された M I D I 楽音データの波形データと歌声データの波形を重ね合わせるによりミキシングを行うことも可能である。

【0053】

D T M (desk top music) 等のシーケンサでは音声波形 (wav形式) のデータを扱えるのは一般的である。上記のように音声波形としてまとめてしまえば D T M 等のシーケンサにおいて、音声波形として取り込むことが可能になり、M I D I の楽音とのミキシング処理自体もシーケンサにより行うことが可能である。

【0054】

一般に M I D I 音源 9 はそのクロック等の違いにより音源の種類により再生される楽音が長い場合にわずかながらズレを生じることが知られている。ズレ補正部 14 ではこのズレを補正するために M I D I 音源 9 の種類にあわせてズレ補正データ 15 内に予め用意された閾値を歌声生成部 7 において歌声を生成する際の時間データに対して掛け合わせることににより補正を行う。

【0055】

このズレ補正データ 15 は歌声生成部 7 が動作している C P U や O S (operating system) などの環境と M I D I 音源の種類の組み合わせによって決まるが、それ以外にオペレータの指示によりこの閾値を変更することも可能である。

【0056】

なお、歌声情報に関して、演奏データに歌詞が含まれている場合を説明したが、これには限られず、演奏データに歌詞が含まれない場合に任意の歌詞、例えば「ら」や「ぼん」等を自動生成し、又はオペレータにより入力し、歌詞の対象とする演奏データ (トラック、チャンネル) を、トラック選択部、歌詞付与部を介して選択して歌詞を割り振るようにしてもよい。

【0057】

図 5 に図 1 に示す歌声合成装置の全体動作をフローチャートで示す。

【0058】

先ず M I D I ファイルの演奏データ 1 を入力する (ステップ S 1)。次に演奏データ 1 を解析し、楽譜データ 4 を作成する (ステップ S 2、S 3)。次にオペレータに問い合わせオペレータの設定処理 (例えば、歌詞の対象とするトラックやチャンネルの指定、ミュート又は音量調整すべきトラック又はチャンネルの指定、wav の作成指示、D T M への取込指示等) を行う (ステップ S 4)。なおオ

ペレータが設定しなかった部分についてはデフォルトが後続処理で使用される。

【0059】

次に、作成した楽譜データに基づき、歌詞を対象とするトラック又はチャンネルの演奏データに割り振って歌声情報6を作成する（ステップS5、S6）。

【0060】

次に上述したタイミングのズレ補正閾値を取得し（ステップS7）、歌声生成部7において歌声情報6から歌声を生成する際の時間データに対して掛け合わせることにより補正を行って、音声波形（歌声波形8）を作成する。

【0061】

次に、MIDI制御データ16を参照して、ミュートすべきトラック、チャンネル又は音量調整すべきトラック、チャンネルがあるかチェックし（ステップ9）該当するMIDIトラック、チャンネルについては対応する処理をする（ステップS10）。代表的には、歌詞の対象とした演奏データ（MIDIトラック、チャンネル）は再生されないか、歌声に比べ小さな音量で再生されるよう音量調整処理される。

【0062】

次に、MIDIからwav形式の作成が指示されているかチェックする（ステップS11）。指示されてなければ、MIDI再生をスタートさせ（ステップS13）、歌声波形8と伴奏波形10との同期を取りながらミキシングする（ステップS17）。

【0063】

MIDIからwav形式の作成が指示されているときは、伴奏波形10を作成した（ステップS14）後、DTMへの取込が指示されているかチェックする（ステップS15）。指示されていれば歌声波形8と共に伴奏波形10をDTMに引き渡す。指示されてなければ歌声波形8と伴奏波形10を重ね合わせる（ステップS16）。

【0064】

ステップS13又はS16の後、D/A変換器、アンプ、スピーカを含むサウンドシステム（図示せず）を介して歌声に伴奏の付いた音響信号を出力する（ス

テップ S 17)。

【0065】

なお、ステップ S 12、S 13 を通って S 17 に進む処理は、代表的には逐次的に実行される。すなわち、MIDI の再生スタートを合図に、順次、リアルタイムでミキシングの実行とサウンドシステムによる音再生が行われる。これに対し、ステップ S 8 からステップ S 14、S 16 を経てステップ S 17 に至る処理の場合、代表的には、いったん（予め）歌声と伴奏音の波形を作成し、重ね合わせてミキシングし、その結果を保存した後に、楽曲のサウンド再生要求に応じて音再生が行われる。

【0066】

以上説明した歌声合成機能は例えば、ロボット装置に搭載される。

【0067】

以下、一構成例として示す 2 足歩行タイプのロボット装置は、住環境その他の日常生活上の様々な場面における人的活動を支援する実用ロボットであり、内部状態（怒り、悲しみ、喜び、楽しみ等）に応じて行動できるほか、人間が行う基本的な動作を表出できるエンターテインメントロボットである。

【0068】

図 6 に示すように、ロボット装置 60 は、体幹部ユニット 62 の所定の位置に頭部ユニット 63 が連結されると共に、左右 2 つの腕部ユニット 64 R/L と、左右 2 つの脚部ユニット 65 R/L が連結されて構成されている（ただし、R 及び L の各々は、右及び左の各々を示す接尾辞である。以下において同じ。）。

【0069】

このロボット装置 60 が具備する関節自由度構成を図 7 に模式的に示す。頭部ユニット 63 を支持する首関節は、首関節ヨー軸 101 と、首関節ピッチ軸 102 と、首関節ロール軸 103 という 3 自由度を有している。

【0070】

また、上肢を構成する各々の腕部ユニット 64 R/L は、肩関節ピッチ軸 107 と、肩関節ロール軸 108 と、上腕ヨー軸 109 と、肘関節ピッチ軸 110 と、前腕ヨー軸 111 と、手首関節ピッチ軸 112 と、手首関節ロール軸 113

と、手部 114 とで構成される。手部 114 は、実際には、複数本の指を含む多関節・多自由度構造体である。ただし、手部 114 の動作は、ロボット装置 60 の姿勢制御や歩行制御に対する寄与や影響が少ないので、本明細書ではゼロ自由度と仮定する。したがって、各腕部は 7 自由度を有するとする。

【0071】

また、体幹部ユニット 62 は、体幹ピッチ軸 104 と、体幹ロール軸 105 と、体幹ヨー軸 106 という 3 自由度を有する。

【0072】

また、下肢を構成する各々の脚部ユニット 65 R/L は、股関節ヨー軸 115 と、股関節ピッチ軸 116 と、股関節ロール軸 117 と、膝関節ピッチ軸 118 と、足首関節ピッチ軸 119 と、足首関節ロール軸 120 と、足部 121 とで構成される。本明細書中では、股関節ピッチ軸 116 と股関節ロール軸 117 の交点は、ロボット装置 60 の股関節位置を定義する。人体の足部 121 は、実際には多関節・多自由度の足底を含んだ構造体であるが、ロボット装置 60 の足底は、ゼロ自由度とする。したがって、各脚部は、6 自由度で構成される。

【0073】

以上を総括すれば、ロボット装置 60 全体としては、合計で $3 + 7 \times 2 + 3 + 6 \times 2 = 32$ 自由度を有することになる。ただし、エンターテインメント向けのロボット装置 60 が必ずしも 32 自由度に限定されるわけではない。設計・制作上の制約条件や要求仕様等に応じて、自由度すなわち関節数を適宜増減することができるというまでもない。

【0074】

上述したようなロボット装置 60 がもつ各自由度は、実際にはアクチュエータを用いて実装される。外観上で余分な膨らみを排してヒトの自然体形状に近似させること、2 足歩行という不安定構造体に対して姿勢制御を行うことなどの要請から、アクチュエータは小型かつ軽量であることが好ましい。また、アクチュエータは、ギア直結型でかつサーボ制御系をワンチップ化してモータユニット内に搭載したタイプの小型 AC サーボ・アクチュエータで構成することがより好ましい。

【0075】

図8には、ロボット装置60の制御システム構成を模式的に示している。図8に示すように、制御システムは、ユーザ入力などに動的に反応して情緒判断や感情表現を司る思考制御モジュール200と、アクチュエータ350の駆動などロボット装置60の全身協調運動を制御する運動制御モジュール300とで構成される。

【0076】

思考制御モジュール200は、情緒判断や感情表現に関する演算処理を実行するCPU (Central Processing Unit) 211や、RAM (Random Access Memory) 212、ROM (Read only Memory) 213、及び、外部記憶装置（ハード・ディスク・ドライブなど）214で構成される、モジュール内で自己完結した処理を行うことができる、独立駆動型の情報処理装置である。

【0077】

この思考制御モジュール200は、画像入力装置251から入力される画像データや音声入力装置252から入力される音声データなど、外界からの刺激などに従って、ロボット装置60の現在の感情や意思を決定する。ここで、画像入力装置251は、例えばCCD (Charge Coupled Device) カメラを複数備えており、また、音声入力装置252は、例えばマイクロホンを複数備えている。

【0078】

また、思考制御モジュール200は、意思決定に基づいた動作又は行動シーケンス、すなわち四肢の運動を実行するように、運動制御モジュール300に対して指令を発行する。

【0079】

一方の運動制御モジュール300は、ロボット装置60の全身協調運動を制御するCPU 311や、RAM 312、ROM 313、及び外部記憶装置（ハード・ディスク・ドライブなど）314で構成される、モジュール内で自己完結した処理を行うことができる、独立駆動型の情報処理装置である。外部記憶装置314には、例えば、オフラインで算出された歩行パターンや目標とするZMP軌道、その他の行動計画を蓄積することができる。ここで、ZMPとは、歩行中の床

反力によるモーメントがゼロとなる床面上の点のことであり、また、ZMP軌道とは、例えばロボット装置60の歩行動作期間中にZMPが動く軌跡を意味する。なお、ZMPの概念並びにZMPを歩行ロボットの安定度判別規範に適用する点については、Miomir Vukobratovic 著“LEGGED LOCOMOTION ROBOTS”（加藤一郎外著『歩行ロボットと人工の足』（日刊工業新聞社））に記載されている。

【0080】

運動制御モジュール300には、図8に示したロボット装置60の全身に分散するそれぞれの関節自由度を実現するアクチュエータ350、体幹部ユニット62の姿勢や傾斜を計測する姿勢センサ351、左右の足底の離床又は着床を検出する接地確認センサ352、353、バッテリーなどの電源を管理する電源制御装置354などの各種の装置が、バス・インターフェース（I/F）301経由で接続されている。ここで、姿勢センサ351は、例えば加速度センサとジャイロ・センサの組み合わせによって構成され、接地確認センサ352、353は、近接センサ又はマイクロ・スイッチなどで構成される。

【0081】

思考制御モジュール200と運動制御モジュール300は、共通のプラットフォーム上で構築され、両者間はバス・インターフェース201、301を介して相互接続されている。

【0082】

運動制御モジュール300では、思考制御モジュール200から指示された行動を体現すべく、各アクチュエータ350による全身協調運動を制御する。すなわち、CPU311は、思考制御モジュール200から指示された行動に応じた動作パターンを外部記憶装置314から取り出し、又は、内部的に動作パターンを生成する。そして、CPU311は、指定された動作パターンに従って、足部運動、ZMP軌道、体幹運動、上肢運動、腰部水平位置及び高さなどを設定するとともに、これらの設定内容に従った動作を指示する指令値を各アクチュエータ350に転送する。

【0083】

また、CPU311は、姿勢センサ351の出力信号によりロボット装置60

の体幹部ユニット 6 2 の姿勢や傾きを検出するとともに、各接地確認センサ 3 5 2, 3 5 3 の出力信号により各脚部ユニット 6 5 R / L が遊脚又は立脚のいずれの状態であるかを検出することによって、ロボット装置 6 0 の全身協調運動を適応的に制御することができる。

【 0 0 8 4 】

また、CPU 3 1 1 は、ZMP 位置が常に ZMP 安定領域の中心に向かうように、ロボット装置 6 0 の姿勢や動作を制御する。

【 0 0 8 5 】

さらに、運動制御モジュール 3 0 0 は、思考制御モジュール 2 0 0 において決定された意思通りの行動がどの程度発現されたか、すなわち処理の状況を、思考制御モジュール 2 0 0 に返すようになっている。

【 0 0 8 6 】

このようにしてロボット装置 6 0 は、制御プログラムに基づいて自己及び周囲の状況を判断し、自律的に行動することができる。

【 0 0 8 7 】

このロボット装置 6 0 において、上述した歌声合成機能をインプリメントしたプログラム（データを含む）は例えば思考制御モジュール 2 0 0 の ROM 2 1 3 に置かれる。この場合、歌声合成プログラムの実行は思考制御モジュール 2 0 0 の CPU 2 1 1 により行われる。

【 0 0 8 8 】

このようなロボット装置に上記歌声合成機能を組み込むことにより、伴奏に合わせて歌うロボットとしての表現能力が新たに獲得され、エンターテインメント性が広がり、人間との親密性が深められる。

【 0 0 8 9 】

なお、本発明は上述した実施の形態のみに限定されるものではなく、本発明の要旨を逸脱しない範囲において種々の変更が可能であることは勿論である。

【 0 0 9 0 】

例えば、本件出願人が先に提案した特願 2 0 0 2 - 7 3 3 8 5 の明細書及び図面に記載の音声合成方法及び装置等に用いられる歌声合成部及び波形生成部に対

応した歌声生成部 7 に使用可能な歌声情報を例示しているが、この他種々の歌声生成部を用いることができ、この場合、各種の歌声生成部によって歌声生成に必要とされる情報を含むような歌声情報を、上記演奏データから生成するようにすればよいことは勿論である。また、演奏データは、MIDI データに限定されず、種々の規格の演奏データを使用可能である。

【0091】

【発明の効果】

以上詳細に説明したように、本発明に係る歌声合成方法及び装置によれば、演奏データを音の高さ、長さ、歌詞の音楽情報として解析し、解析された音楽情報に基づき、歌声を生成するとともに、上記演奏データに基づき、上記歌声の伴奏としての楽音を生成することにより、MIDI データに代表されるような演奏データ（楽器制御データ）から楽音の再生のみならず楽音を伴奏として歌詞を歌い上げることができる。したがって、従来、楽器の音のみにより表現していた音楽の作成や再生において特別な情報を加えることがなく歌声の合成を行うことによりその音楽表現は格段に向上する。

【0092】

また、本発明に係るプログラムは、本発明の歌声合成機能をコンピュータに実行させるものであり、本発明に係る記録媒体は、このプログラムが記録されたコンピュータ読み取り可能なものである。

【0093】

本発明に係るプログラム及び記録媒体によれば、演奏データを音の高さ、長さ、歌詞の音楽情報として解析し、解析された音楽情報に基づき、歌声を生成するとともに、上記演奏データに基づき、上記歌声の伴奏としての楽音を生成することにより、演奏データ（楽器制御データ）から楽音の再生のみならず楽音を伴奏とした歌唱が可能となる。

【0094】

また、本発明に係るロボット装置は本発明の歌声合成機能を実現する。すなわち、本発明のロボット装置によれば、供給された入力情報に基づいて動作を行う自律型のロボット装置において、入力された演奏データを音の高さ、長さ、歌詞

の音楽情報として解析し、解析された音楽情報に基づき、歌声を生成するとともに、上記演奏データに基づき、上記歌声の伴奏としての楽音を生成することにより、MIDIデータに代表されるような演奏データ（楽器制御データ）から楽音の再生のみならず楽音を伴奏として歌詞を歌い上げることができる。したがって、ロボット装置の表現能力が向上し、エンターテインメント性を高めることができると共に、人間との親密性を深めることができる。

【図面の簡単な説明】

【図 1】

本実施の形態における歌声合成装置のシステム構成を説明するブロック図である。

【図 2】

解析結果の楽譜情報の例を示す図である。

【図 3】

歌声情報の例を示す図である。

【図 4】

歌声生成部の構成例を説明するブロック図である。

【図 5】

本実施の形態における歌声合成装置の動作を説明するフローチャートである。

【図 6】

本実施の形態におけるロボット装置の外観構成を示す斜視図である。

【図 7】

同ロボット装置の自由度構成モデルを模式的に示す図である。

【図 8】

同ロボット装置のシステム構成を示すブロック図である。

【符号の説明】

2 演奏データ解析部、5 歌詞付与部、7 歌声生成部、9 MIDI音源
11 ミキシング部、12 MIDI制御部、60 ロボット装置、211 CPU、213 ROM

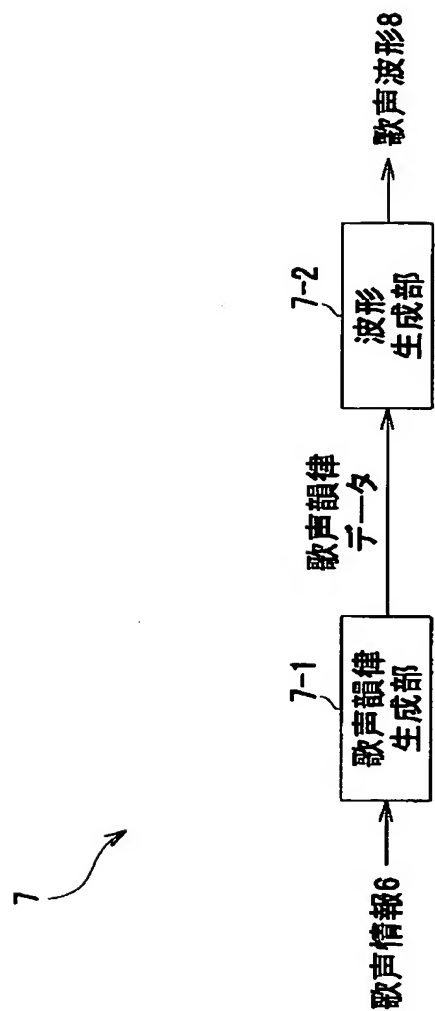
【図 2】

トラック	チャンネル	時間	タイプ	高さ	長さ	強さ	長さ	コントロールの種類
1	1	5:03:480	∴	-	-	-	-	ビブラート (深さ64 幅64 遅れ50)
1	1	5:03:480	ノート	G4	199	100	あ	
1	1	5:04:000	ノート	F#4	439	108	る	
1	1	5:04:480	ノート	G4	199	100	う	
1	1	6:01:000	ノート	E4	199	90	ひ	
2	1	4:01:480	コントロール	-	-	-	-	Expression (110)
2	1	4:01:480	コントロール	-	-	-	-	ビブラート (深さ64 幅64 遅れ50)
2	1	6:01:480	ノート	G3	199	100	あ	
2	1	6:02:000	ノート	F#3	439	108	る	
2	1	6:02:480	ノート	G3	199	100	う	
2	1	6:03:000	ノート	E3	199	90	ひ	
			∴					

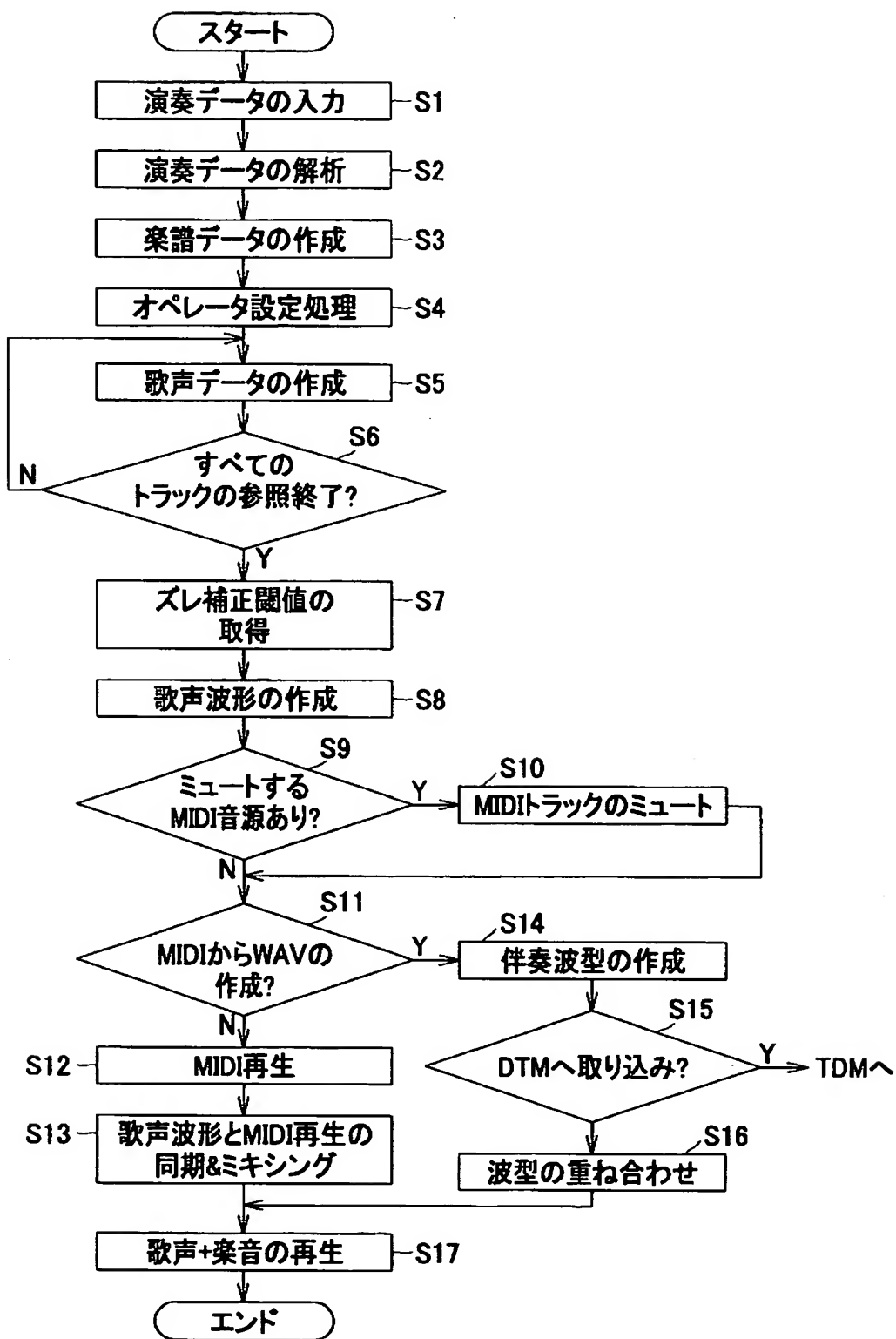
【図 3】

¥song¥	← 歌声データの開始
¥PP,T10673075¥	← 10673075 μ secの休み
¥tdyna 110 649075¥	← 先頭から10673075 μ secの全体の強さ
¥fine-100¥	← ピッチの微調整(MIDIのファインチューンに同じ)
¥vibrato NRPN_dep=64¥	← ビブラート
¥vibrato NRPN_de=50¥	
¥vibrato NRPN_rat=64¥	
¥dyna 100¥	← 音ごとの強弱
¥G4,T288461¥あ	← G4の高さの音で、長さが288461 μ sec、歌詞は「あ」
¥dyna 108¥	
¥Gb4,T288462¥る	
¥dyna 100¥	
¥G4,T288461¥う	
¥dyna 90¥	
¥E4,T219592¥ひ	
¥PP,T1222716¥	
¥dyna 100¥	
¥E4,T144231¥も	
¥dyna 98¥	
¥E4,T144230¥り	
⋮	

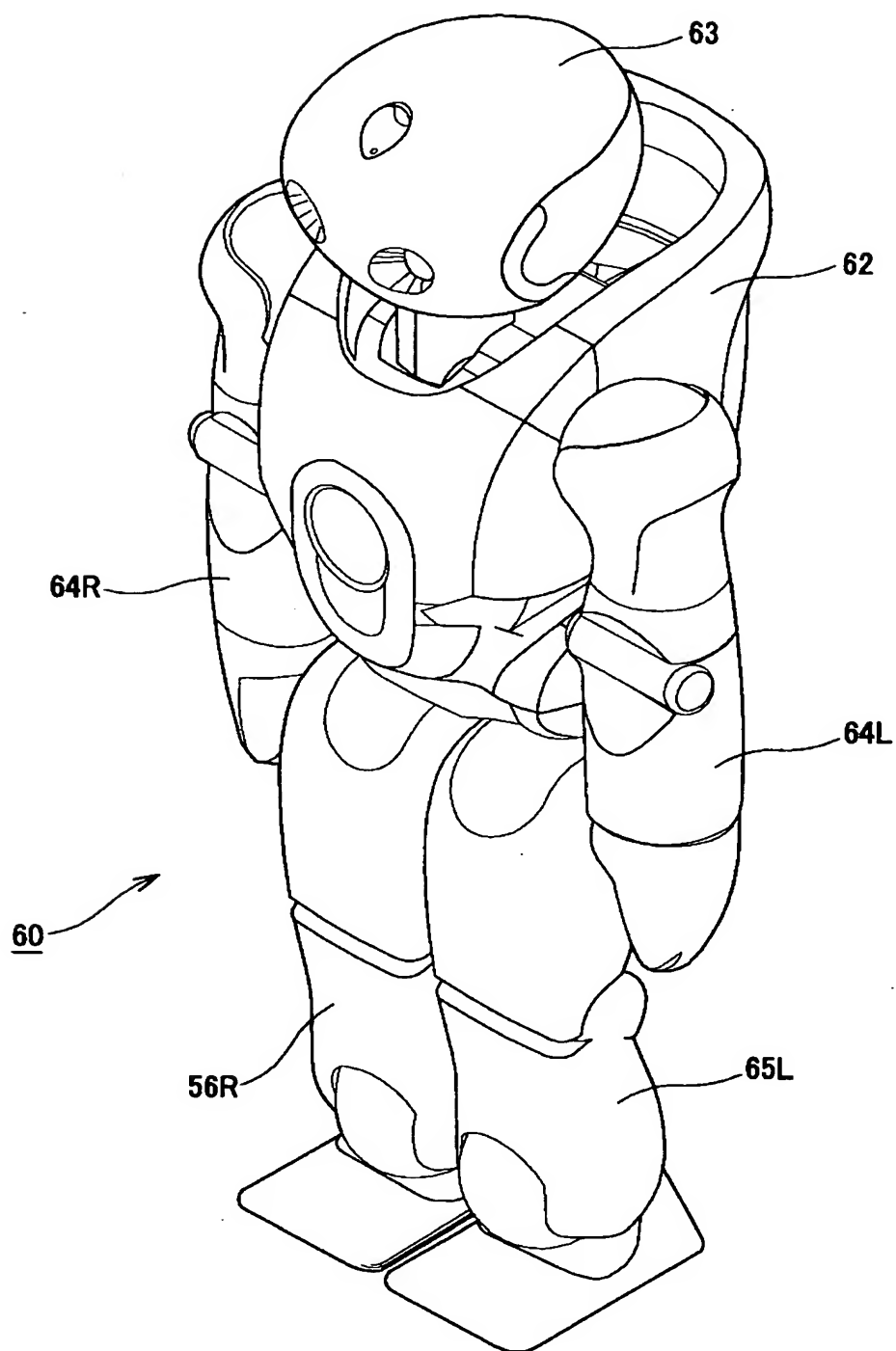
【図 4】



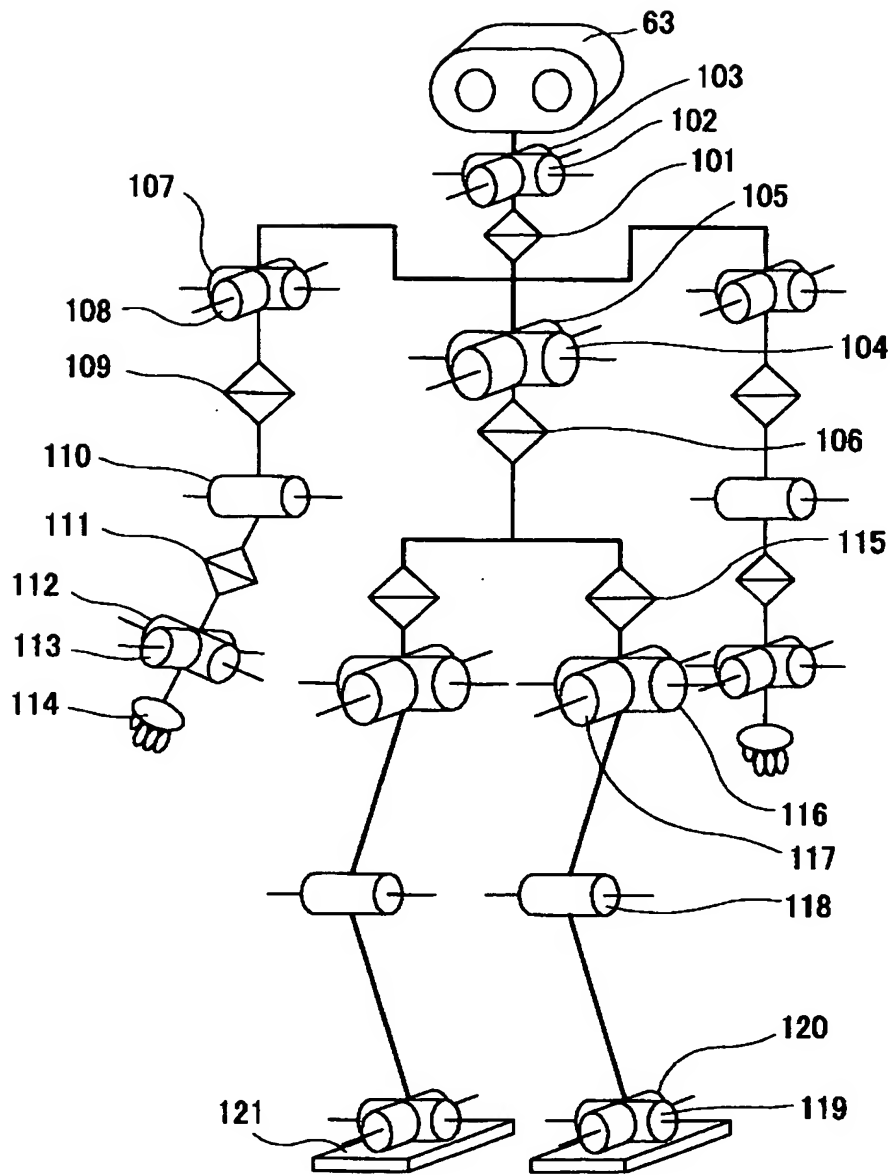
【図 5】



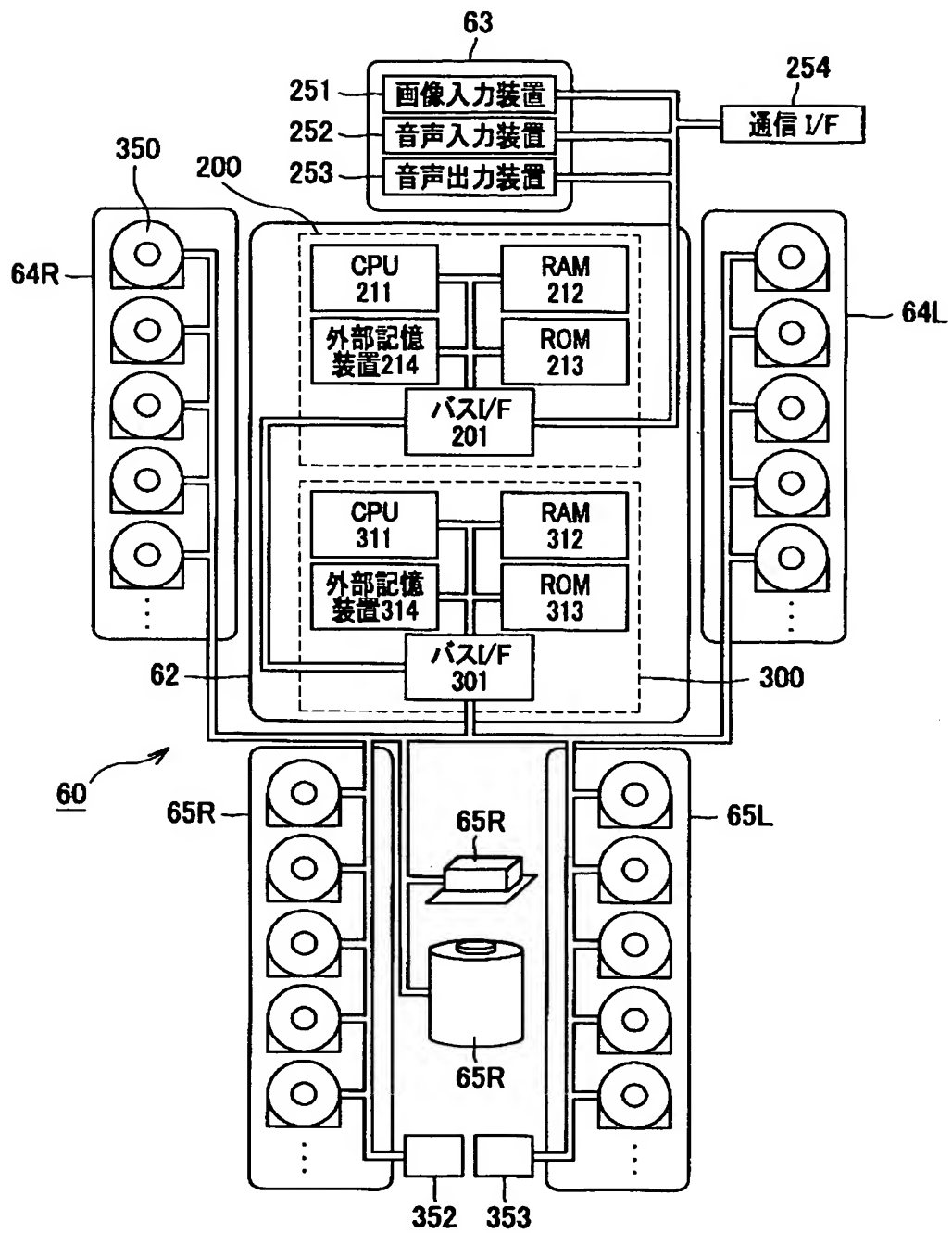
【図 6】



【図 7】



【図 8】



【書類名】 要約書

【要約】

【課題】 M I D I データ等の演奏データを活用して歌声を合成する。

【解決手段】 入力された演奏データを音の高さ、長さ、歌詞の音楽情報として解析する（S 2、S 3）。解析された音楽情報から歌詞を音列に付与して歌声データを作成する（S 5、S 6）。歌声データから歌声の音声波形を作成する（S 7、S 8）。入力された演奏データから楽音の波形を作成する（S 1 4）。歌声に使用した演奏データは楽音の再生に使用しないか再生の音量を抑制することが好ましい。

【選択図】 図 5

特願 2 0 0 3 - 0 7 9 1 5 1

出 願 人 履 歴 情 報

識別番号

[0 0 0 0 0 2 1 8 5]

1. 変更年月日

1 9 9 0 年 8 月 3 0 日

[変更理由]

新規登録

住 所

東京都品川区北品川 6 丁目 7 番 3 5 号

氏 名

ソニー株式会社